



7-15-2009

Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation: A Case Study in Woodford County, Kentucky

Philip B. Mink
Kentucky Archeological Survey

John Ripy
University of Kentucky, jripy@uky.edu

Keiron Bailey
University of Arizona

Ted H. Grossardt
University of Kentucky, tedgrossardt@gmail.com

Follow this and additional works at: https://uknowledge.uky.edu/ktc_facpub



Part of the [Civil and Environmental Engineering Commons](#)

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Mink, Philip B.; Ripy, John; Bailey, Keiron; and Grossardt, Ted H., "Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation: A Case Study in Woodford County, Kentucky" (2009). *Kentucky Transportation Center Faculty and Researcher Publications*. 12.
https://uknowledge.uky.edu/ktc_facpub/12

This Conference Proceeding is brought to you for free and open access by the Transportation at UKnowledge. It has been accepted for inclusion in Kentucky Transportation Center Faculty and Researcher Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation: A Case Study in Woodford County, Kentucky

Notes/Citation Information

Published in the *2009 Proceedings of ESRI Users Conference*, Paper 1495.

The copyright holders have granted the permission for posting the article here.

Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation

Philip B Mink
Kentucky Archeological Survey
1020a Export Street
Lexington, Kentucky 40506-9854
United States
8592578207
philip.mink@uky.edu

Mr. John Ripy
University of Kentucky Transportation Center
176 Raymond Bldg
Lexington, Kentucky 40506-0281
United States
8592577536
jripy@uky.edu

Dr. Keiron Bailey
University of Arizona
Harvill Building Box #2
Department of Geography and Regional Development
Tucson, Arizona 85712
United States
5206121652
kbailey@email.arizona.edu

Dr. Ted Grossardt
Univerisyt of Kentucky Transportation Center
176 Raymond Bldg
Lexington, Kentucky 40506-0281
United States
8592577522
tgrossardt@uky.edu

A paper prepared for the 2009 ESRI International Users Conference
Submitted May 15, 2009
Presentation July 15, 2009

Presentation Title

Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation

Presentation Abstract

Analytic predictive archaeological models can have great utility for state Departments of Transportation, but it is difficult to model the likelihood of prehistoric settlement using geographical proxy predictor variables because of the complexity of how settlement choices were actually made, and the complex interaction between these variables using GIS. In many cases classic statistical modeling approaches require too much data to be useful. This research reports on a preliminary predictive model that combines Spatial Analyst and fuzzy logic modeling to capture expert archaeological knowledge and convert this into predictive surface. A test area was defined in Woodford County, KY and five influencing factors were defined and calculated using ArcMap. Locations were sampled and probabilities estimated using both small and large group structured processes from a range of archeologists that fed an iterative fuzzy logic induction process. An output probability function was generated to create a predictive decision support layer.

Presentation Will Feature the Following ESRI Software

ArcInfo, Spatial

Predictive Archaeological Modeling using GIS-Based Fuzzy Set Estimation: A Case Study in Woodford County, Kentucky

This paper reports on the development and testing of a GIS/fuzzy-logic based predictive model for an area of the Inner Bluegrass physiographic region in Kentucky. The objective of this research was to develop a spatial decision support system, or SDSS that would allow the Environmental Division of the Kentucky Transportation Cabinet to better spatially estimate the probability of encountering archaeological artifacts of a given progeny and epoch, in this case prehistoric lithic scatters. This information would be useful to most state DOT's at the initial stages of developing preferred corridors for new highways, in that it would aid in developing corridors that minimize the likelihood of encountering, and thus being delayed by, significant archeological finds. Although predictive archaeological modeling is not a new field (1,2), the method reported here is novel in several ways. It represents an integration of geographic variables with expert judgments. This model integrated physiographic predictive factors derived from archeological literature into a geospatial platform, ArcGIS, with expert judgments regarding the interaction effects of the combinations of factors governing the probability of prehistoric settlement placement. The factors used were judged by the archeologists to be the most fundamental, or universal, types of factors, with the expectation that they would form the basis of a robust modeling logic in many regions. Those factors include level ground, the relative proximity to a water supply, and elevations that avoid frequent flooding. Additional regionally-specific factors such as springs and sinkholes were introduced as the model was developed, but the foundational considerations were not altered.

The combination of fuzzy logic based nonlinear system modeling and geospatial information management has not been applied to this domain before. Fuzzy logic techniques are well-established in fields such as systems engineering (3) and biological systems modeling (e.g. 4). They are used to model input-output relationships between variables where data is insufficient to support statistical analysis. Fuzzy logic approaches involve categorizing inputs, in this case the variables that affect settlement likelihood, and outputs, in this case probability of discovering artifacts in that location, into small number of discrete classes, and then using neural network or cellular automata algorithms to model the input-output response function based on the available, usually sparse data. For example, the output, probability of encountering prehistoric lithic scatters, was categorized into five classes: very low probability, low, medium, high and very high. Numerical value correspondence was defined for each of these categories. Similarly, inputs were also categorized: for example, slope consisted of very low, low, high and very high categories, each of which corresponds to a specific gradient range. This classification of variables is performed by expert groups using verbal descriptors, thus setting the appropriate level of precision and accuracy for the problem set. The output is a database in the fuzzy model builder software that can be explored graphically to determine probabilities for input combinations that were not sampled (5).

Fuzzy logic approaches deliver effective predictive capacity within the domain of their outputs. Fuzzy systems offer advantages for this type of work: in addition to allowing robust input-out models to be built in data-sparse conditions, they effectively handle nonlinear relationships between variables where there are too many interactions to model effectively and comprehensively using standard statistical techniques. Comparisons between classic statistical modeling techniques and fuzzy approaches are not generally appropriate because fuzzy logic is best used in domains where the sample size is a small fraction of the total decision space,

conditions under which statistical models are useless. In this case, the potential sets of combinations of inputs was large and the actual domain of knowledge was small, leading to the selection of fuzzy logic as an appropriate and meaningful modeling approach. One disadvantage of fuzzy systems is that they operate using categorical variables, and although these can be translated into numerical ranges, they cannot be relied on to generate precise estimates for output variables measured on interval or ratio scales. Nevertheless, the logic of a fuzzy-set based model allows for more classes of outputs than the classic Boolean presence/absence dependent variable. Integrating the fuzzy logic model in this way allowed a range of spatial analytic operations to be performed on multiple layers of data and it facilitated comprehensive visualization of final probability distributions across a relatively large geographic area.

Over a two year period, from summer 2006 to 2008, the Kentucky Transportation Cabinet's archaeological team consulted with team members with expertise in nonlinear modeling using fuzzy set estimation methods and geospatial platform development to develop and apply the methodology. In this paper the aims of the modeling process are defined, the selection of the case study area is explained, the process of generating the model is described and the iterative refinement procedure is detailed. Results from the case study area of Woodford are presented, lessons learned during model build are discussed, and the strengths and weaknesses of the method are summarized.

Aims and Scope

The utility of an effective predictive archaeological model is significant for State Departments of Transportation. For new highway builds, or rehabilitation, or for the many other infrastructure development processes that involve the exercise of eminent domain, archaeological inspection is a costly and time-intensive process. Potential legal challenges and EIS mandates require good faith efforts to investigate and dig to locate and identify potential sites. The Kentucky Transportation Cabinet's archaeological team desired a geospatial decision support tool that would help them visually understand the distribution of probabilities of encountering artifacts from a given period over a meaningful area of interest. The tool was also required to provide a reasonable quantitative categorical estimate of probability mapped to specific locations.

With this context in mind, the model described here is designed explicitly to be deployed by archeologists as a first step in the assessment of likelihood for prehistoric artifacts. The model is not intended to be interpreted and employed without considerable archaeological expertise, because it is designed to accommodate professional archeological knowledge in addition to the parameters contained within the model. Its purpose is to allow rapid broad-picture summarization, and to focus archaeological expertise more efficiently on areas of high interest, not to eliminate areas from consideration. The primary utility of the model is to aid in selection of a corridor when there are multiple corridors under consideration, as is often the case in early planning stages. The model can be used to aid in sorting corridors that have the least probability of encountering archaeological sites. It can provide supporting data for automated tools for alignment evaluations similar to the research team's Analytic Minimum Impedance Surface, or AMIS, method (6).

The focus of this model is on prehistoric, open-air lithic scatters, partly because it allowed the use of a fairly robust database for testing model results. While it would undoubtedly be desirable to model only the locations of National Register of Historic Places (NRHP)-eligible sites, such sites are so sparsely located and highly variable in age and type that there is

insufficient data to attempt to model. Significance is associated with factors such as integrity, site type, or the actual age of the site, variables not used in this locational predictive model. This is a model of simple site location, and therefore will most accurately reflect the distribution of the most common site types. Thus, while the model cannot predict site significance, it can provide an estimate of potential site density. Statistically, the corridor with the greatest site density also has the greatest potential for NRHP-eligible sites.

Initially, the appropriate spatial scale of the model was debated. It was considered that statewide application would not be feasible because Kentucky contains a number of physiographic regions that exhibit widely differing landscape characteristics. Modeling rules derived to represent the influence of factors such as distance above water, for example, could not be expected to be consistent across different landforms. The individual physiographic region, within which landscape form, characteristics and variations were predictable and similar, was felt to be a suitable domain. The Inner Bluegrass geological region of central Kentucky, a raised Karst landscape, was considered to be a good candidate. It is an area within which a number of projects are underway, or will soon be, requiring archaeological survey. This area also had considerable historic data concerning archaeological discoveries, essential for model testing and verification. Finally, each member of the archaeological team also possessed a number of years experience working on surveys and sites within this region. This factor was critical because the modeling process sought to capture the expert knowledge of the team, beyond literature reviews, and convert this into the probability surface.

One advantage of using fuzzy logic approaches to system modeling is that it facilitates the capture of this implicit knowledge; that is, system knowledge that participants or respondents possess but are not necessarily capable of articulating though description of formal mathematical relationships between input variables. They may have what they call “gut feelings,” or “intuition,” or “best guess” knowledge that is based on years of experience, observations and analysis in the field. This knowledge cannot be accessed explicitly, but it can be captured and implicitly decomposed using a structured approach to knowledge creation such as multivariable modeling (7,8).

Predictive archaeological modeling

Archaeologists have been attempting to model potential archaeological site locations since the early 1980's, and with the availability of desktop GIS packages these types of studies have increased over the past couple of decades. Previous attempts to model sites have had only limited success and principally focused on inductive methodologies where known sites were correlated to environmental variables (9). These projects employ statistical techniques such as linear regression on existing site databases and modern environmental data to map areas that have either a high or low probability for containing archaeological sites. In North Carolina, nearly 5000 archeological sites and a “wide range of environmental variables” were used as data to build a logistic regression model that sorted the landscape into High, Medium, or Low probabilities (10). The study funded by the Minnesota Department of Transportation Mn/Model used a similar approach and, after three iterations, settled on 44 different environmental variables to use in predicting the location of archeological sites (11).

A more local example of this approach is a project undertaken by the authors for predicting the location of Archaic Period archaeological sites in Henderson County, Kentucky in 2001, with the goal of examining the influence of representing sites as a singular point or as

polygons for GIS locational modeling (12). It successfully developed a robust GLM model for predicting potential site locations (approximately 75% of site locations were correctly predicted within a high probability area that covered roughly 30% of the study area) and demonstrated the benefits of using polygon site boundary locations for archaeological modeling. However, the model's performance was limited by the lack of linearity in the relationship between archaeological sites and environmental variables.

This complexity stems from many factors. First, the locations may have been in use intermittently or continuously from several hundred years ago up to 12,000 years ago. Many kinds of disturbances to sites can occur over that time span, so sites that are correctly modeled as having been inhabited may nonetheless yield no evidence of the people, or artifacts, either because it too faint to see or has been obliterated by later prehistoric or modern human activity.

The second major issue is attempting to understand and reconstruct the culture and decision-making of humans from an entirely different culture. For example, previous archaeological research indicates that the most recent Native American farmers would tend to place villages near good farmland and water, and prior to them hunters and gathers would have had smaller campsites located in an area with good visibility and in good habitat for deer, turkey and other animals. Both groups produced quarries and other resource extractions sites such as chert outcrops. However, much less is known about their decisions guiding the location of religious, ceremonial, or burial sites. Thus, any predictive modeling attempt should draw most strongly on those relationships that are best understood, and capitalize on the existing body of archeological knowledge. Because not all relationships and decisions are yet fully understood, a modeling attempt cannot expect to decisively 'predict' the locations of all artifacts, but should yield predictive outputs that are consistent with archeologists' understandings, and, of course, the evidence in the field. Consequently, constructing predictive rules should rely more on relationships regarding water access and quality, suitability for habitation, visibility for hunting, etc.

Steps to Model Build

The first step was to determine significant factors that influence likelihood of settlement during the period in question and to arrange these into discrete categories amenable to a fuzzy set mapping. Following a literature review, a structured brainstorming and group classification session was hosted with the archaeology team. Five key predictor factors were identified by archaeological team members and the output variable, probability, was defined as shown in Table 1.

Table 1. Variables matrix

Degrees Slope	Minutes Walk to Water	Minutes Walk to Confluence	Distance Above Water in Feet	Stream Rank in Strahler Order	Probability Scale
1 = (VL <= 5)	L <= 2	L <= 10	VL <= 10	L = 1	VL <= 1.2
2 = (L, 5 - 10)	M = 2 - 4	H = >10	L = 10 - 25	M = 2 - 3	L = 1.2 to 2.5
3 = (H, 11 - 20)	H = >5		H = 25 - 60	H = >3	M = 2.7 to 3.2
VH = >20			VH = >60		H = 3.3 to 3.9
					VH = >3.9

Each input variable was divided into classes that made sense to the archaeologists. These classifications are shown in Table 1. The number of classes for each variable represents the judgment of the team regarding the number of discrete categories that influence human behaviors, and therefore that control settlement and habitation patterns for each. These were based on a combination of the teams' knowledge of archaeological literature and a discussion held during the meeting. It is important to note that, while these parameters were generally identified as fundamental to habitation decisions of prehistoric peoples, this particular classification system of distances and heights might not have validity outside of the defined study area. It was developed specifically to match the team's expert knowledge of the topography, hydrology, and physical geography of the Woodford County study area. For example, for the height above water categories, three significant classes were defined. To establish category boundaries, attention was paid to local terrace topography in the vicinity of the Ohio and Kentucky Rivers. These parameters were also restricted to those that were derivable from existing GIS data, so that the model could be developed using available Kentucky data without demanding significant new data gathering.

The first factor identified was **Slope in Degrees** of the surface. The basic argument here is that it is difficult to maintain a habitation on ground that is too steep, and increasingly difficult to impossible as slope reaches certain values. Slope was given four categories: VL = ≤ 5 , L = 5-10, H = 10-20, and VH = 20+ degrees of slope.

The second interactive factor identified was **Minutes walk to nearest walkable water** (using Tobler's algorithm for computing walking time). This factor was divided into three categories: L ≤ 2 , M = 2-4, and H = 4+ minutes. Again, the general presumption was that access to water promotes the likelihood of settlements, but this is conditioned by other factors in the model.

The third interactive factor identified was **Minutes walk to nearest walkable confluence on streams with a Strahler order of 3 or higher (13)**. This factor was divided into only two categories: L ≤ 10 , H = 10+ minutes. The inclusion of this factor was meant to account for the significance of confluences of larger streams as an attractive factor for habitation location.

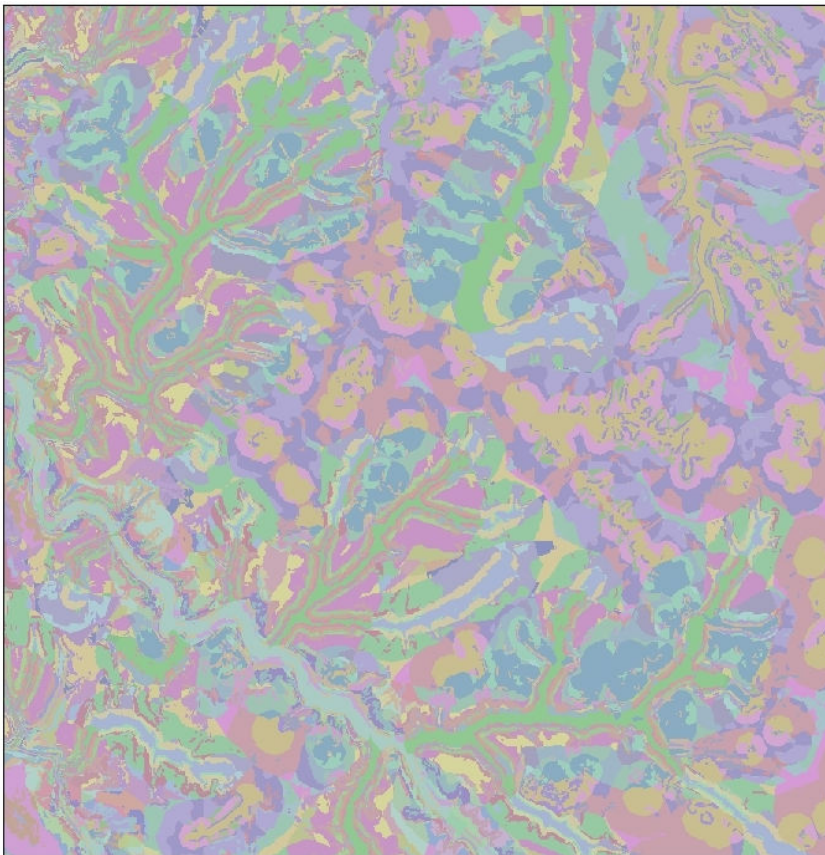
The fourth interactive factor identified was **Elevation difference to nearest walkable water (not direct line) in feet**. This was divided into four categories: VL ≤ 10 , L = 10-25, H = 25-60, and VH = 60+ feet. This factor helps account for the risk of flooding at very low elevations and the attractiveness of various landforms high above river bottoms.

The fifth interactive factor identified was the **Strahler order of the streams**. Although the Strahler system allows for stream orders as high as 10, the factor was divided into three categories: L, M, and H. For purposes of the model, two different definitions of category membership were used. Under one definition set, L = 1, M = 2,3, and H = 4+. Under the second definition set, L = 1,2, M = 3, and H = 4. This factor helps mediate the relative impact of distances to water by the size and reliability of the water source, as indicated through the Strahler order.

After being defined in this way, several of these inputs needed to be converted into spatial dependence functions that represented accessibility. Horizontal distance from water, for example, was converted into minutes walking time. This variable was computed geospatially, using a custom non-istropic cost-distance function created in ArcGIS to measure the distance to the nearest watercourse, adjusted for topographical variation along that least horizontal distance path, and divided by average walking speed to produce time to water. The Strahler stream order model, while seemingly conceptually simple, required the derivation of a set of rules for automating the designation of streams consistently in a landscape heavily dissected by all manner of watercourses. “Neighboring” was defined as the most proximate watercourse. One problem with this was that while some points were located near only one watercourse, others were located within almost equal adjacency of two or three watercourses.

Once coverages were generated, they were categorized into polygons according to the rules set out above. This yielded six new coverages, each with anywhere from two to four polygon categories. These coverages were then spatially intersected to produce all the possible combinations of the classifications. This potential number is given by multiplying the number of categories for each factor by the number of categories for the next factor and so on, yielding a total of $4 \times 3 \times 2 \times 4 \times 3 = 288$ possible unique spatial landscape categories. In fact the actual intersection yielded about 180 unique landscape categories within the boundaries of the testing county (Figure 1).

Figure 1. Color-coded GIS surface illustrating 180 unique spatial landscape categories resulting from combinations of feature coding categories: Woodford County, KY.



A GIS output surface showing the topography of the region was used to elicit probability estimations from the team. At the structured meeting, the archaeological team was asked to evaluate the probability of encountering an artifact in a set of sample locations (cells in the raster surface) under consideration on a scale of 1 (extremely unlikely) to 5 (extremely likely). The valuations of the team were computed using the arithmetic mean. A default value of ‘Moderate’ was assigned to all locations where insufficient knowledge or conditions existed to indicate either higher or lower likelihoods of artifacts. This position was adopted because it was judged more accurate to incorporate no bias in either direction about unknown areas. The input valuations to the cell were interrogated using the GIS query function and recorded. This produced an input-output mapping function and was recorded in a spreadsheet. The process was repeated for all fifty sample points. Toward the end of this process, some points were found to represent duplicates of earlier points and if there was divergence from the previous probability value, the reasons were discussed. When all fifty points had been captured, they were input into the *FuzzyKnowledgeBuilder* software. The software’s cellular automata function was invoked and a knowledge base was built. This final knowledge base contained a complete functional mapping of outputs (probability) across the full range of all six input parameters. The process is similar to the generation of the community knowledge base used by the authors in CAVE protocols for visual evaluation, with the exception that in this case the output is probability (14,15). The output from the software was interrogated across the full range of each of the input parameters and each corresponding input-output mapping was recorded in a spreadsheet lookup table. The final lookup table contained all 288 possible landscape combinations. Table 2 shows a small sample of the output.

Table 2. Sample input-output lookup table extracted from knowledge base, diagramming the relationship between various coded combinations of landform factors and the probability, on a scale of 1-5, of encountering prehistoric artifacts

Slope	Mins to Water	Mins to Confluence	Dist Above Water	Stream Rank	Probability
1	1	1	1	1	3
1	1	1	1	2	2
1	1	1	1	3	2
1	1	1	2	1	3
1	1	1	2	2	3
1	1	1	2	3	5

The researchers also explored the robustness of the input by presenting a strategic subset of the landform samples to a statewide meeting of professional archeologists. Using the GIS, they explained the logic of the predictive system. Then, for each of about 30 sites, the relevant landform properties were supplied and discussed, and then the group scored each site on a scale of 1-5, corresponding to the five categories of likelihood specified earlier. Approximately 70 professionals entered their individual judgments anonymously through the use of electronic keypads, and the means of these scores were reviewed by, and compared to, the initial scores developed by the research team. Based on the scores and discussion at the meeting, adjustments were made to specific cell combinations as appropriate.

At a subsequent structured meeting, the GIS surfaces showing the physical and environmental variables and the probability surface was displayed in tandem with the spreadsheet, and a careful evaluation was made for consistency of judgment. This evaluation resulted in recategorizing approximately 30% of the total observations. Appropriate manual adjustments were made to these individual values in the knowledge base, preserving the remainder of the modeled values.

Finally, these tabular probability values were converted into a continuous geographic probability surface in ArcGIS. The input parameters for each cell in the landscape were converted to outputs using the lookup table data applied using the ArcGIS platform's "Reclassify" function for raster data. The output, probability, was color coded across the range of 1 through 5, showing all probabilities ranging from very low to very high across the Woodford County study area. The team inspected the surfaces and the distribution. The purpose of this inspection was to ensure that spatial patterns at the larger scale matched the expert knowledge of the archaeological team. This led to a round of backward iteration, i.e. an adjustment of the rules and the knowledge base.

The use of the GIS distribution to evaluate coherence allowed team members to adjudicate spatial mismatches between the modeled probabilities and their assessments. This phase was valuable, because in certain higher-level topographic areas far from water, probability showed as low or very low when team members felt it should be much higher. This effect recurred on many higher plateau, and it was considered sufficiently regular to point to a systemic omission in the input variables. The team decided that the presence of sinkholes in the Karst landscape was responsible for this apparent mismatch. Sinkholes were considered to be locations around which the probability of prehistoric settlement was higher, regardless of the other factors.

Consequently, sinkholes were treated as independent, additive adjustments to the final probability model i.e. as a second spatial layer that could be overlain onto the original probability surface, and arithmetically combined using raster addition, or subtraction, logic. The presence of sinkholes was considered to modify the probability of settlement, all other factors held equal, depending on net distance from the sinkhole. Beyond a certain distance from source, the sinkhole enhancement effect diminished to zero. To geocode this probability adjustment, a known sinkhole location layer was registered and added, and a two-zone buffer was created around each sinkhole location. Within 100 feet of the sinkhole, probability was assigned a +2 rating, and from 100 to 500 feet it was assigned a +1 rating. A raster addition operation was performed on the sinkhole adjustment and the original probability layers to generate a final probability layer. The effect of this additional layer was to double the potential number of unique combinations of landscape features from 288 to 576 (Table 3). Also, it increased the numeric probability scores up to 7 (5+2), however scores from 5-7 all remained classified as 'very high probability.'

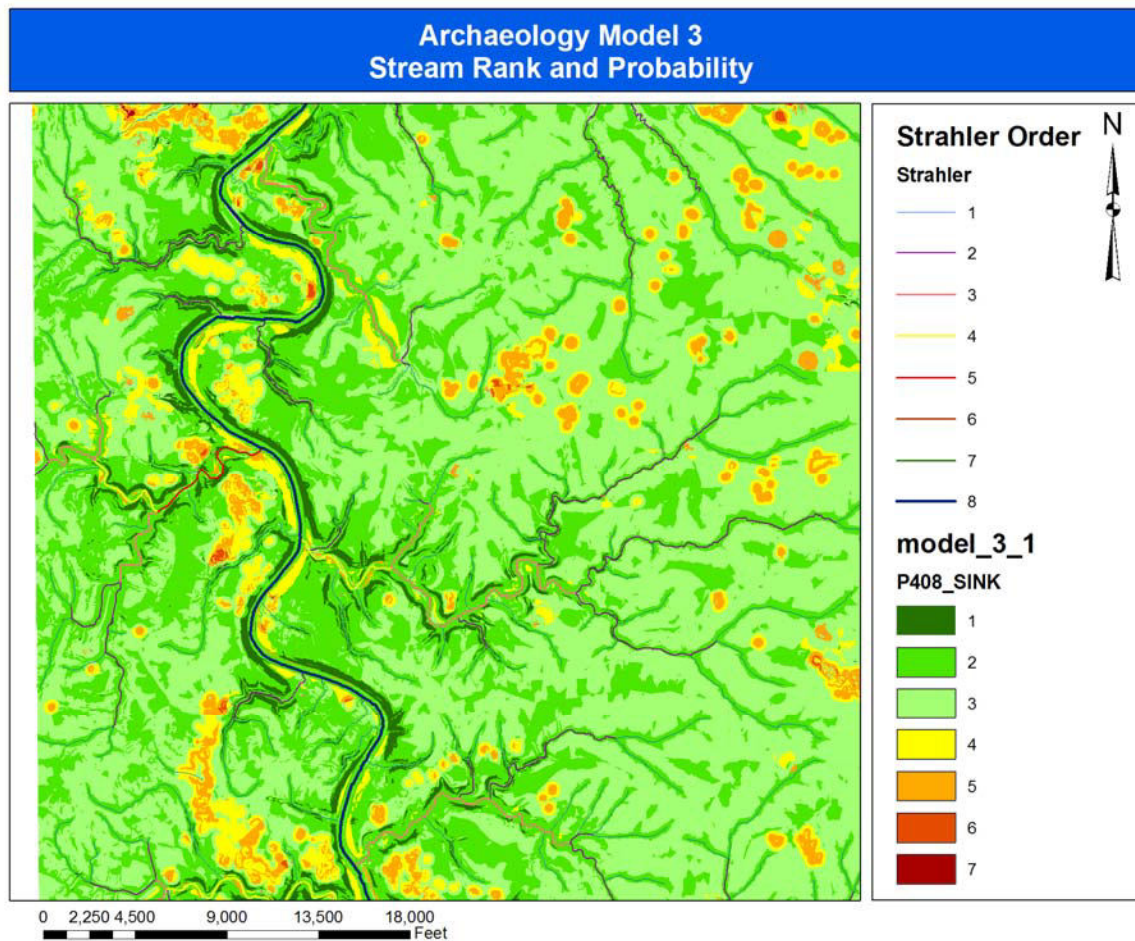
Table 3. Modification of original probability model (see Table 2) to accommodate proximity to sinkholes as an additive factor.

Initial Model Probability	Distance to Spring	Resulting Change in Probability	Final Probability Value
3	< 100'	+2	5
2	100-500'	+1	3
2	>500'	0	2
3	100-500'	+1	4
3	>500'	0	3
5	<100'	+2	7

The parameter ranges were evaluated by the archaeological team and a number of adjustments were proposed. These were incorporated into the input-output lookup tables. Because the effects of certain classifications were not easy to gauge from the lookup tables, it was proposed to develop a series of models that provided coverage across input element variation. This process was constrained by group discussion and agreement to six specifications.

Six models were then built that each incorporated slight changes in the categories shown in Table 1. For example, “Minutes Walk to Water” coding might have been reclassified as L<=3, M= 4-5, and H = 6+ for one alternate model. A typical result, this one for “Model 3” is shown in Figure 2.

Figure 2. Typical archaeological probability mapping output for Woodford County, KY, where increasing likelihood of encountering prehistoric artifacts is indicated by increasingly red color codes 1-7. This surface is derived from the application of the rating 'rules' to the 180 unique surface categories shown in Figure 1.



Results of verification

Results for known artifact locations for Woodford County were extracted from the state database. These were independently plotted by one of the team members, separately from the modeling team. Half of the points (about 50) were used to compare with the predicted model outputs and discrepancies were tabulated. To ensure integrity for the verification process, the team members responsible for modeling did not come into contact with the verification database at any point. The performance of the models is shown below, in Table 4.

Discussion

Avoiding overprediction errors was considered a more important design objective than dealing with the effects of underprediction. The model was calibrated to load uncertainty into the probability class 3, or moderate. Therefore, these cells displaying this value can be

interpreted either to mean moderate probability of encountering artifacts, or as cells about which little is known. The rationale for this approach was to focus the predictive capacity of the model on the more extreme probabilities, both high and low. Engineering the systemic uncertainty into this region was intended to increase the discrimination of the model at the margins, particularly in the extreme probability categories 1 (very improbable) and 5 (likely to encounter artifacts).

The results validate this design characteristic. Each model performs slightly differently in absolute terms across its range, but the basic pattern of the results is similar. The models show high performance for all categories other than medium category (Very Low, Low, High, Very High). These ratio values range from a low of .77 to a high of .98, where “1” represents 100% efficiency (Table 4). By comparison, the Minnesota Archeological Predictive Model shows efficiency ratio values for their predicted “High” probability zones of between .49 and .92, and lower efficiencies of .28-.89 for High/Medium categories. However, reaching these predictive ratios required the use of 44 variables in logistic regression equations (16).

Table 4: Comparison of the technical “efficiency” of six model iterations in correctly predicting the relative density of artifacts in a landscape. The efficiency ratio is computed as 1- (% of area/% of sites) so that a perfectly efficient model would yield a ratio of 1, and a completely random model would yield a ratio around 0. This ratio was developed by Kvamme as a way of comparing the efficiency of various archeological predictive models (9).

The major problem created by the concentration of uncertainty in the moderate probability category is the difficulty this creates for interpretation. Another problem is the degree to which the chosen predictor variables capture all the potential influences on settlement location. The five variables chosen clearly cannot account for all the possible reasons for choosing settlement location. Models of this type can never be 100% efficient, because of the considerable simplification of complex, historical decisionmaking processes forced by this logic. Another problem was the variable definitions. The boundaries between categories were based on existing literature and team experience, and adjusted as part of the overall model adjustment process.

Conclusions

This county-scale model shows promising

VL (1)	Sink 1	0.9713
L (2)	Sink 1	0.8041
M (3)	Sink 1	0.3722
H (4)	Sink 1	0.9283
VH (5)	Sink 1	0.9242
VL (1)	Sink 2	0.9831
L (2)	Sink 2	0.7871
M (3)	Sink 2	0.3748
H (4)	Sink 2	0.9234
VH (5)	Sink 2	0.9316
VL (1)	Sink 3	0.9832
L (2)	Sink 3	0.7759
M (3)	Sink 3	0.3749
H (4)	Sink 3	0.9249
VH (5)	Sink 3	0.9412
VL (1)	Sink 4	0.9892
L (2)	Sink 4	0.7817
M (3)	Sink 4	0.3751
H (4)	Sink 4	0.9242
VH (5)	Sink 4	0.9297
VL (1)	Sink 5	0.9838
L (2)	Sink 5	0.7920
M (3)	Sink 5	0.3913
H (4)	Sink 5	0.9110
VH (5)	Sink 5	0.9218
VL (1)	Sink 6	0.9840
L (2)	Sink 6	0.7785
M (3)	Sink 6	0.3880
H (4)	Sink 6	0.9162
VH (5)	Sink 6	0.9333

preliminary results. The major success of this approach is shown by the capacity of the models to meet or exceed the measured efficiency of existing statistical models using a much more modest set of input variables. Given the limited time and human resource expended on constructing the model, and the efficiency with which the prediction surface was built, the team considers this to be a successful result. These initial models are somewhat more efficient at predicting archeological resources than are existing linear regression archaeological models, despite our cautious approach of assigning most 'unknowns' into the 'medium' probability category.

The application area is currently limited to a county. Moreover, the team does not consider it feasible to use this approach to build a monolithic model beyond the scale of the physiographic region because the sensitivity of the input parameters, and therefore the performance of the model, would be diminished. However, for applications over large areas, such as entire states, it would be possible to mosaic a series of regional models, effectively providing much larger coverage, using ArcGIS functionality. The team is now working in collaboration with other Kentucky Transportation Cabinet to extend the modeling approach to the entire surrounding physiographic region, and to enhance the accuracy of the current model by further iterative refinement.

REFERENCES

- (1) Kvamme, K. 1983 Computer Processing Techniques for Regional Modeling of Archaeological Site Locations. *Advances in Computer Archaeology* 1:26-52.
- (2) 1984 Models of Prehistoric Site Location Near Piñon Canyon, Colorado. In *Papers of the Philmont Conference on the Archaeology of Northeastern New Mexico*, edited by C. J. Condie, pp. 347-370. Proceedings of the New Mexico Archaeological Council 6(1). Albuquerque
- (3) Zadeh, L. Fuzzy Sets, *Information and Control Vol. 8 No. 3*, 1965, pp.338-353.
- (4) Ridgley M, and R. Ruitenbeek. *Optimization of Economic Policies and Investment Projects Using a Fuzzy-Logic Based Cost-Effectiveness Model of Coral Reef Quality: Empirical Results for Montego Bay, Jamaica. Coral Reefs Vol. 18 No. 4*, 1999, pp.381-392.
- (5) Bailey, K., Grossardt T., and Brumm, J. 2001. "Towards Structured Public Involvement in Highway Design: A Comparative Study of Visualization Methods and Preference Modeling using CAVE (Casewise Visual Evaluation)," *Journal of Geographic Information and Decision Analysis* 5: 1-15
- (6) Grossardt, T., K. Bailey, and J. Brumm. Structured Public Involvement: Problems and Prospects for Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1858*, TRB, National Research Council, Washington, D.C., 2003, pp. 95-102.
- (7) Ridgley M, Fernandes L. Multiple criteria analysis integrates economic, ecological and social objectives for coral reef managers. *Coral Reefs* 1999;18(4):393-402.
- (8) Meesters E, Bak R, Westmacott S, Ridgley M, Dollar S. A fuzzy logic model to predict coral reef development under nutrient and sediment stress. *Conservation Biology* 1998;12(5):957-65
- (9) Kvamme, Kenneth L. "There and Back Again: Revisiting Archaeological Locational Modeling." In *GIS and Archaeological Site Location Modeling*, 3-38. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2006.
- (10) Seibel, Scott. Archeological Predictive Modeling. *Government Engineering*. Sept-Oct 2006. pp. 35-37.
- (11) Hobbs, E. and Nawrocki, T. Minnesota Department of Transportation. Chapter 6: Archaeological and Environmental Variables: A Predictive Model of Precontact Archeological Site Location for the State of Minnesota. <http://www.mnmodel.dot.state.mn.us/chapters/chapter6.htm> accessed July 30, 2008

- (12) Mink, Philip B. II, B. Jo Stokes, and David Pollack. "Points vs. Polygons: A Test Case Using a Statewide Geographic Information System." In *GIS and Archaeological Site Location Modeling*, 219-240. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2006.
- (13) Tobler, Waldo. *Three Presentations On Geographic Analysis and Modeling*. Santa Barbara, CA: National Center for Geographic Information and Analysis, University of California, February 1993. http://www.ncgia.ucsb.edu/Publications/Tech_Reports/93/93-1.PDF.
- (14) Bailey, K., T. Grossardt, and M. Pride-Wells. Community Design of a Light Rail Transit-Oriented District using CAVE (Casewise Visual Evaluation). *Socio Economic Planning Sciences*, Vol. 41, No. 3, 2007, pp. 235-254.
- (15) Bailey, K., and T. Grossardt. Structured Public Involvement in Context-Sensitive Noise Wall Design using Casewise Visual Evaluation (CAVE). In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1984, TRB, National Research Council, Washington, D.C., 2006, pp. 112-120.
- (16) Hobbs, E., Johnson, C. and Gibbon, G. Chapter 7: Model Development and Evaluation in *A Predictive Model of Precontact Archeological Site Location for the State of Minnesota*. Table 7.6 at <http://www.mnmodel.dot.state.mn.us/chapters/chapter7.htm#tbl76>, accessed November 7, 2008.

